Annales Universitatis Paedagogicae Cracoviensis
Studia Naturae, 10: XX–XX, 2025
ISSN 2543-8832
DOI: 10.24917/25438832.10.x

Kirolos Eskandar
Helwan University, Faculty of Medicine and Surgery – Egypt;
email: kiroloss.eskandar@gmail.com

Bioinformatics meets AI: transforming big data into breakthroughs in biotechnology

Abstract

The convergence of bioinformatics and artificial intelligence (AI) is revolutionising the biotechnology landscape, transforming biological big data into actionable insights and groundbreaking innovations. Bioinformatics, with its ability to manage and analyse massive datasets from genomics, proteomics, and systems biology, faces challenges of complexity and scale. AI's advanced techniques – machine learning, deep learning, and natural language processing-provide unprecedented tools for deciphering patterns, making predictions, and driving automation. This synergy has catalysed remarkable progress in precision medicine, drug discovery, gene editing, and synthetic biology, heralding a new era of data-driven breakthroughs. This review explores the foundations, transformative applications, and recent breakthroughs at the intersection of bioinformatics and AI, while addressing challenges and envisioning a future where interdisciplinary collaboration unlocks the full potential of this powerful partnership.

Keywords: artificial intelligence, big data, bioinformatics, biotechnology innovation, computational biology
Received: [2025.03.30]
Accepted: [2025.04.22]

Introduction

Bioinformatics, a dynamic interdisciplinary field, integrates biology, computer science, and statistics to process and interpret the staggering volumes of data generated by modern biological research. This data, often derived from next-generation sequencing, proteomics, metabolomics, and structural biology, holds the potential to unlock profound insights into life processes (Pereira et al., 2020). However, the complexity, diversity, and scale of biological big data pose significant challenges in storage, analysis, and interpretation. Traditional computational tools, while invaluable, are increasingly inadequate to handle this deluge of information efficiently (Rawat, Yadav, 2021).

Artificial Intelligence (AI) has emerged as a transformative technology capable of addressing these challenges. AI encompasses a range of computational methods, including machine learning (ML), deep learning (DL), and natural language processing (NLP), all designed to emulate aspects of human intelligence (Collins et al., 2021). These methods excel at identifying patterns in data, predicting outcomes, and automating tasks. In the field of biotechnology, AI has demonstrated remarkable success, from predicting protein structures with unprecedented accuracy to accelerating drug discovery pipelines and deciphering genomic variants associated with disease (Larabi-Marie-Sainte et al., 2019; Jumper et al., 2021).

The intersection of bioinformatics and AI represents a revolutionary paradigm shift. This convergence leverages AI's computational power to enhance the capabilities of bioinformatics, addressing critical bottlenecks such as data integration, noise reduction, and multidimensional analysis. By combining these two disciplines, researchers can derive actionable insights from complex datasets, enabling advancements in precision medicine, synthetic biology, and personalized therapeutics (Jamialahmadi et al., 2024). For instance, AI-enhanced bioinformatics pipelines are instrumental in decoding cancer genomes, optimising gene-editing tools like CRISPR, and understanding microbiome-host interactions (Alm, 2024).

The aim of this review is to explore the transformative potential of combining AI and bioinformatics in biotechnology. By examining their foundational principles, recent breakthroughs, and practical applications, this article aims to provide a comprehensive understanding of how these fields are converging to drive innovation. The review is structured to first outline the key concepts and technologies underpinning bioinformatics and AI, followed by an exploration of their synergistic applications. Additionally, it highlights recent breakthroughs, discusses technical and ethical challenges, and offers insights into future directions for research and collaboration. This discussion is intended not only for researchers and practitioners in computational biology but also for a broader audience interested in the profound implications of AI-powered bioinformatics on healthcare, agriculture, and environmental sciences. By illuminating this vibrant interdisciplinary frontier, this review underscores the immense opportunities for innovation at the nexus of bioinformatics and AI.

Methodology

Search strategy

A systematic literature review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure methodological transparency and reproducibility. Comprehensive searches were performed across prominent scientific databases, including PubMed, Google Scholar, Scopus, and Web of Science. The search spanned publications from January 2019 to February 2025.

The search strategy was meticulously designed to capture the breadth and depth of the topic. The keywords and Medical Subject Headings (MeSH) terms used included: “Artificial Intelligence”, “Bioinformatics”, ‘Big Data’, ‘Biotechnology’, and “Computational Biology”. Boolean operators (AND, OR) were employed to refine and combine search terms, enhancing precision and comprehensiveness. The search was restricted to peer-reviewed articles published in English.

Study selection

The inclusion criteria for selecting studies were as follows:

- Language: studies published in English.
- Focus: articles explicitly discussing AI applications in bioinformatics or biotechnology, particularly regarding big data analysis and innovation.
- Study type: original research, systematic reviews, and meta-analyses.
- Publication date: articles published between 2019 and 2025.

Exclusion criteria included:

1. Non-peer-reviewed sources (e.g., editorials, commentaries, conference abstracts).
2. Studies lacking a direct focus on AI, bioinformatics, and biotechnology.
3. Studies with insufficient methodological rigor (e.g., lacking clear data analysis methods).
4. Papers with unresolved conflicts in quality assessment (e.g., studies with unclear methodologies and conflicting results).

Screening process

The initial database search yielded 264 articles. After removing duplicates using EndNote and Mendeley, 175 unique records remained. These were subjected to a two-stage screening process:

1. Title and abstract screening: the titles and abstracts of all articles were reviewed for relevance. Articles failing to meet the inclusion criteria were excluded.
2. Full-text screening: full-text versions of the remaining articles were assessed against the inclusion and exclusion criteria.

After this screening process, 92 studies met all criteria and were included in the final review.

Data extraction

Data were systematically extracted using a standardised form. Key data points collected included:

- Study characteristics: author(s), year of publication, study design, and region of origin.
- AI applications: specific AI methods or algorithms used (e.g., machine learning, deep learning, natural language processing) and their roles in bioinformatics and biotechnology.
- Big data integration: description of datasets analysed (e.g., omics data, structural biology data, clinical datasets).
- Outcomes and limitations: study findings, reported challenges, and limitations, including algorithmic bias and dataset quality.

The data extraction process was independently conducted by two reviewers, with discrepancies resolved through discussion to ensure accuracy.

Quality assessment

The Joanna Briggs Institute (JBI) Critical Appraisal Tools were employed to evaluate the quality of the selected studies. Criteria such as methodological rigor, risk of bias, and result validity were assessed. All 92 studies included in this review met high-quality standards, ensuring robust methodologies, rigorous validation, and minimal risk of bias. The studies were assessed based on the clarity of their research design, data collection methods, statistical analyses, and overall reliability. Since all included studies demonstrated strong methodological rigor, no studies were excluded due to quality concerns. This ensures that the synthesis and conclusions of this review are built upon reliable, peer-reviewed, and well-supported research.

Data synthesis

A mixed narrative and descriptive synthesis approach was employed to analyse and present findings from the included studies. Data were organized thematically to highlight key areas of AI application, such as big data analysis, precision medicine, drug discovery, and genomics.

To enhance clarity and objectivity, a quantitative layer was added to the synthesis by categorising all 92 included studies based on AI techniques used, domains of application, and publication year. These trends were analysed using frequency counts and percentage distributions. The synthesized data are visually summarised through tables and figures presented in the Results section. These visual elements support the identification of dominant AI methods, domain-specific trends, and temporal publication patterns.

Due to heterogeneity in study designs, AI architectures, and research objectives, a formal statistical meta-analysis was not feasible. However, this structured descriptive approach enables a comprehensive and transparent exploration of AI's role in bioinformatics. To provide a transparent overview of the study selection process, a PRISMA flow diagram is included (Fig. 1), detailing the number of records identified, screened, and included in the review, along with reasons for exclusion at each stage.

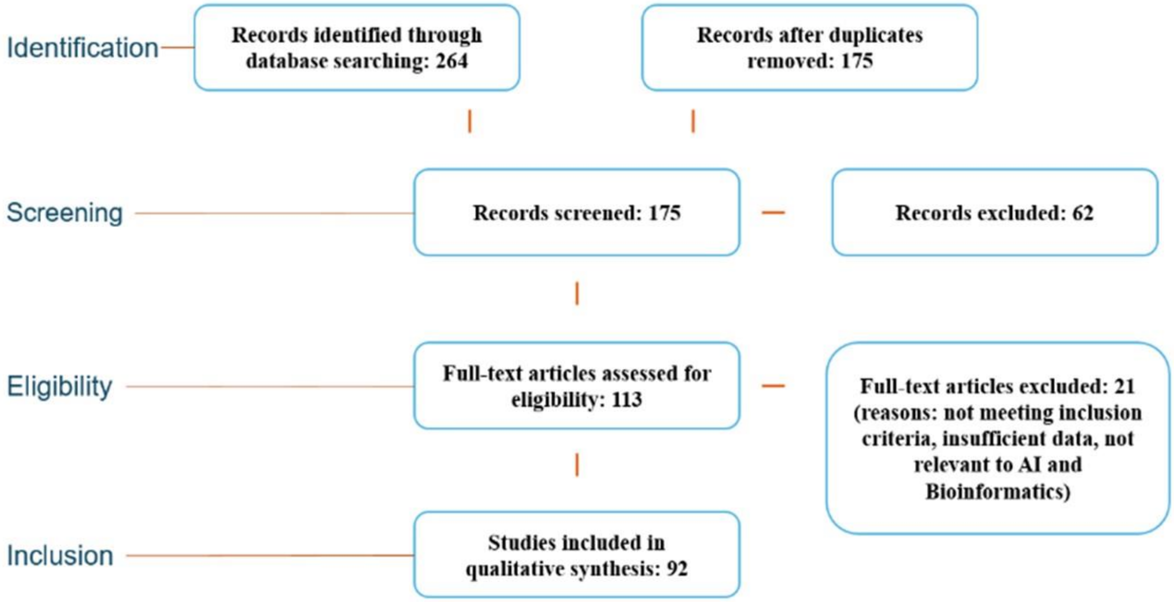

Fig. 1. The PRISMA flow diagram

Results of meta-analysis

This section presents the quantitative synthesis of 92 peer-reviewed studies analysed in this review. The goal is to illuminate trends in the adoption of artificial intelligence (AI) across bioinformatics applications by categorising AI methods used, domains of application, and publication timelines.

1. AI techniques used (Tab. 1; Fig. 2)

Across all 92 studies, AI techniques were classified into primary categories. Some studies used multiple methods.

Tab. 1. The AI techniques used in Frequencies and Percentages

AI Technique	Frequency	Percentage (%)
Machine Learning (ML)	43	46.7
Deep Learning (DL)	35	38.0
Natural Language Processing (NLP)	13	14.1
Quantum AI / Quantum-Assisted ML	9	9.8
Federated / Privacy-Preserving AI	4	4.3
Evolutionary Algorithms	3	3.3
Explainable AI (XAI)	5	5.4

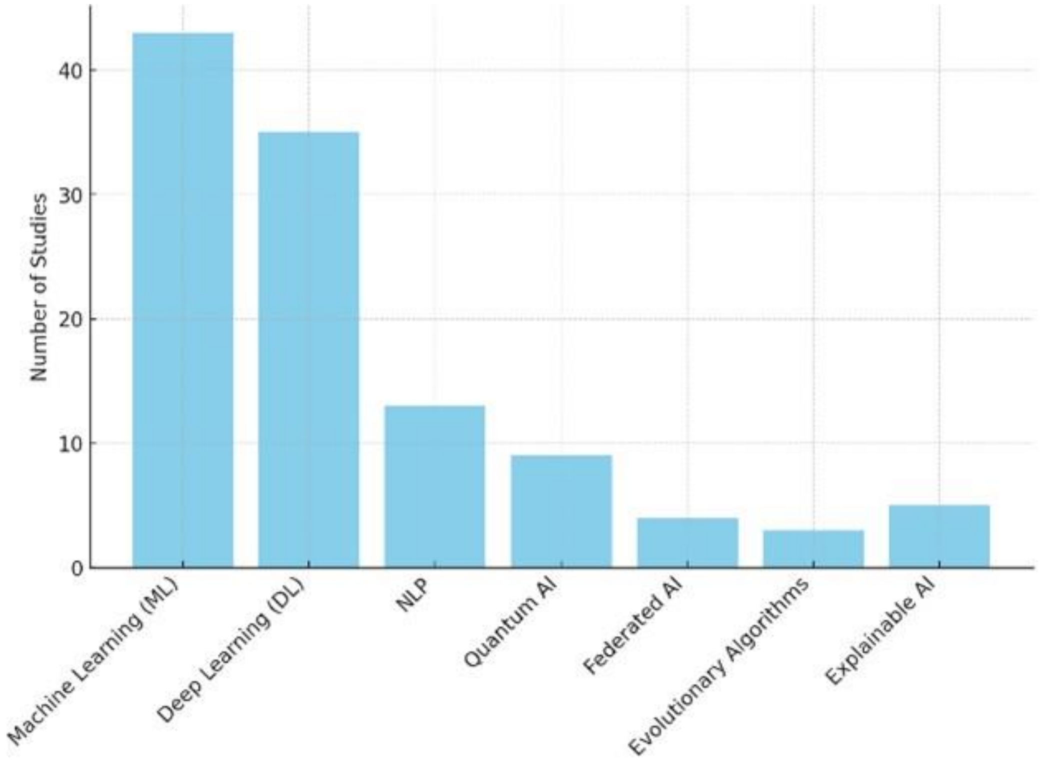

Fig. 2. The frequency of AI techniques in bioinformatics studies

2. Bioinformatics domains of application (Tab. 2; Fig. 3)

The studies were mapped to their bioinformatics domains based on the context and datasets used. Genomics remains the leading domain for AI applications, with drug discovery and multiomics close behind.

Tab. 2. The domains used in references included in this article

Domain	Frequency	Percentage (%)
Genomics and Genome Analysis	31	33.7
Multiomics Integration	16	17.4
Proteomics and Structural Biology	14	15.2
Drug Discovery and Development	21	22.8
Precision/ Personalised Medicine	11	12.0
Microbiome/ Metagenomics	4	4.3
Network Biology/ Systems Biology	6	6.5

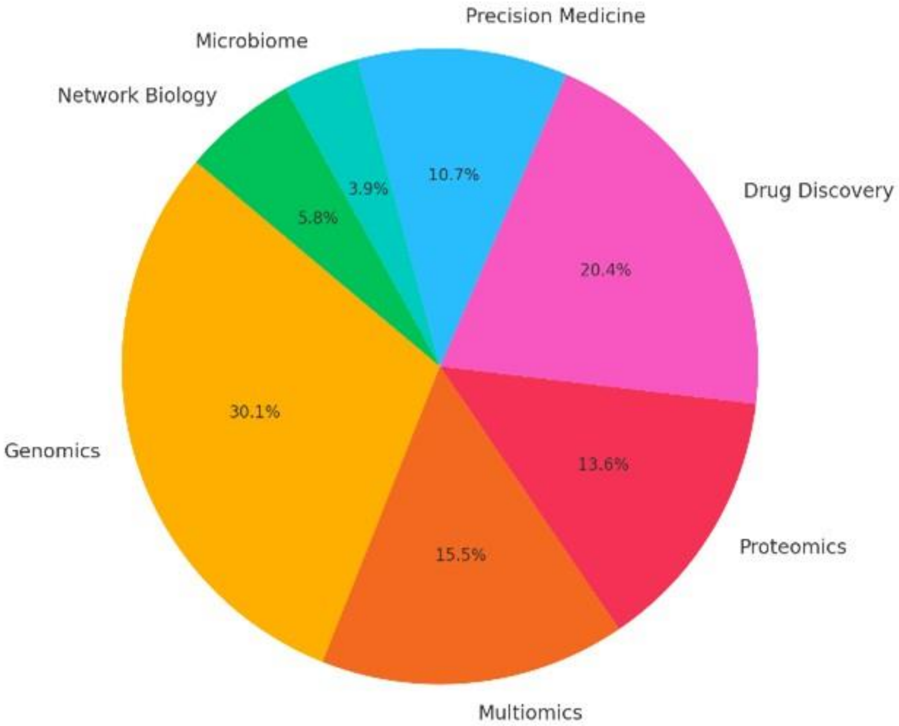

Fig. 3. The distribution of bioinformatics domains

Publication year analysis shows a sharp acceleration in the last three years (Tab. 3).

Tab. 3. The Number of publications used in this article according to their years of publication

Year	Number of publications
2019	2
2020	1
2021	4
2022	9
2023	24
2024	38
2025	14 (as of early 2025)

3. Cross-domain and cross-tech trends

- AlphaFold and structural DL models dominate proteomics research from 2021 onward.
- NLP methods (e.g., BioBERT, literature mining) are increasingly applied in systems biology and clinical data mining.
- Quantum-assisted AI has grown in visibility but remains exploratory.
- Privacy-preserving techniques (e.g., federated learning) appear primarily in studies from 2023–2025, signalling emerging interest in ethical AI integration.

Discussion

The foundations of bioinformatics and AI in biotechnology

Bioinformatics serves as the cornerstone for the analysis and interpretation of complex biological data, enabling researchers to decode the information embedded in DNA, RNA, and proteins. Advances in sequencing technologies, such as next-generation sequencing (NGS) and single-cell sequencing, generate massive datasets with unparalleled detail, covering entire genomes, transcriptomes, and proteomes (Satam et al., 2023). These omics technologies have revolutionized our understanding of biological systems but also introduced significant computational challenges. For example, integrating and analysing multi-omics datasets is complicated by data heterogeneity, noise, and sheer scale (Zhao et al., 2022). Specialised bioinformatics tools are required to manage and interpret this data, from assembling genomic sequences to modelling metabolic networks and visualising complex interactions between cellular components (Aradhya et al., 2023).

Simultaneously, artificial intelligence has emerged as an indispensable tool in addressing these computational demands. At its core, AI encompasses machine learning (ML), a branch of algorithms designed to identify patterns and relationships in data without explicit programming. Among ML techniques, deep learning has shown remarkable efficacy in extracting insights from high-dimensional data (Datta et al., 2024). Neural networks, particularly convolutional and recurrent architectures, are now extensively applied in bioinformatics to predict protein folding, identify genetic variants associated with diseases, and analyse gene expression patterns (Jamialahmadi et al., 2024). These methods enable tasks such as unsupervised clustering of transcriptomic data, prediction of protein-protein interactions, and real-time data-driven diagnostics (Quazi, 2022).

The integration of AI into bioinformatics workflows significantly enhances their efficiency and accuracy. AI algorithms enable the interpretation of complex datasets by overcoming computational bottlenecks, automating repetitive processes, and enhancing scalability (Ayyagiri et al., 2024). For instance, deep learning models trained on large datasets of known genetic mutations and phenotypic outcomes can predict the pathogenicity of novel mutations with impressive accuracy (Brandes et al., 2023). Furthermore, natural language processing (NLP) algorithms have been employed to extract and synthesize knowledge from the vast corpus of scientific literature, offering researchers insights into emerging discoveries across the life sciences (Aradhya et al., 2023; Spurney et al., 2021).

This convergence of AI and bioinformatics transforms biotechnology by enabling researchers to tackle questions that were previously intractable. AI-powered approaches are drive precision medicine by integrating genomic data with patient-specific information to tailor therapies (Carini, Seyhan, 2024). They also accelerate the discovery of novel drug candidates, optimising CRISPR gene-editing systems, and elucidating the molecular basis of complex diseases like cancer and neurodegenerative disorders (Nojadeh et al., 2023). The ability of AI to augment bioinformatics pipelines ensures that biological big data can be translated into meaningful insights, advancing our understanding of life processes and empowering innovations in medicine and agriculture (Huo, Wang, 2024).

Big data in bioinformatics: a growing challenge

Biological big data has grown exponentially due to advancements in high-throughput sequencing, structural biology, and clinical research, yielding datasets of unprecedented scale and complexity. This phenomenon is aptly described by the “3Vs”: volume, variety, and velocity (Munawar et al., 2022). Volume captures the immense size of datasets, exemplified by terabytes generated in a single next-generation sequencing (NGS) run. Variety underscores the diversity of data types, from genomic sequences to proteomic profiles, medical imaging, and patient records. Velocity refers to the rapid pace of data generation, driven by technologies like single-cell sequencing and real-time biosensors (Yang et al., 2023; Vitorino, 2024).

Managing this deluge of data is a formidable challenge. Storing such vast datasets requires scalable solutions, often leveraging cloud computing infrastructure to address capacity and access demands (Al-Kateeb, Abdullah, 2024). Data integration is equally complex, necessitating harmonisation across disparate sources, such as multi-omics datasets, which vary in formats, scales, and processing techniques. Effective integration enables researchers to identify meaningful correlations and build holistic models of biological systems (Kumar et al., 2023).

Interpretation is particularly challenging due to the inherent noise, heterogeneity, and high dimensionality of biological data. Sophisticated machine learning techniques, including neural networks and clustering algorithms, are essential for deriving insights. For example, AI models have been employed to detect disease-associated patterns in genomic data, providing key insights into conditions like cancer and rare genetic disorders (Badrulhisham et al., 2023; Ng et al., 2023).

Computational systems biology has emerged as a solution to address these challenges by focusing on the networks and interactions underlying biological phenomena. Network-based models, including dynamic biomarkers, provide powerful tools to study diseases as systems-level perturbations rather than isolated molecular events (Tang et al., 2022). These advancements have enabled breakthroughs in precision medicine, drug discovery, and translational research, paving the way for transformative applications in biotechnology (Zitnik et al., 2024).

Key AI Techniques in Bioinformatics

The integration of Artificial Intelligence (AI) into bioinformatics has led to transformative advances in the field, with specific AI techniques enabling significant breakthroughs in data analysis, structure prediction, and pattern recognition (Singh et al., 2023). Among these, machine learning (ML) methods such as clustering, classification, and regression have been pivotal in analysing complex datasets from genomics, proteomics, and other omics studies (Arjmand et al., 2022). For example, AI-powered ML techniques are extensively used to identify disease biomarkers by uncovering hidden patterns in large-scale datasets (Chen et al., 2025).

Deep learning, a subset of AI, has revolutionised structural biology, particularly in protein folding and prediction. AlphaFold, developed by DeepMind, stands as a landmark achievement by predicting protein structures with remarkable accuracy directly from amino acid sequences (Yang et al., 2023). Building on this foundation, AlphaFold-Multimer has expanded these capabilities to accurately model protein-protein interactions, facilitating breakthroughs in structural biology and rational drug design. Its ability to predict the structures of protein complexes has accelerated the discovery of protein-target interactions, aiding in antibody design, enzyme engineering, and therapeutic protein formulation (Uzoeto et al., 2024). This advancement has been instrumental in drug discovery, where understanding protein-ligand binding and receptor interactions is crucial for developing novel therapeutics. In structural biology, AlphaFold-Multimer has provided new insights into macromolecular assemblies, enabling researchers to explore previously intractable protein complexes with high confidence (Varadi et al., 2022).

Additional innovations, like RoseTTAFold and CollabFold, have introduced computational efficiencies and broadened accessibility to protein modelling (Nussinov et al., 2022). Natural Language Processing (NLP) is another key AI application, enabling bioinformatics to extract insights from biomedical literature and clinical records. These methods aid in mining critical information for drug-target interactions and patient stratification in personalized medicine. Platforms like BioBERT have significantly improved the extraction of context-specific biological insights (Q. Chen et al., 2021).

Evolutionary algorithms further contribute by optimising molecular designs and refining models in systems biology. They leverage principles of evolutionary biology to develop new drugs and synthesize novel compounds, showcasing AI's adaptability to various biological challenges (Vora et al., 2023).

Transformative applications at the intersection

The integration of AI and bioinformatics has catalysed transformative advances across several domains in biotechnology. Precision medicine, drug discovery, genomics, synthetic biology, and microbiome research have all benefited from the synergies of these two fields (Mohseni, Ghorbani, 2024).

In precision medicine, AI-powered bioinformatics has enhanced the identification of biomarkers, improved patient stratification, and enabled predictive analytics for treatment outcomes. AI-driven algorithms analyse omics data and clinical records to identify disease subtypes and optimize therapeutic strategies (Liao et al., 2023). For instance, machine learning has been instrumental in linking genetic variants with disease risks, thereby facilitating personalized care (Umapathy et al., 2023).

Drug discovery and development have been accelerated by AI’s ability to analyse large molecular datasets. AI models excel at predicting drug-target interactions, enabling faster identification of viable drug candidates (Visan, Negut, 2024). Applications in virtual screening and drug repurposing are now commonplace, with AI reducing the time and cost of the drug development pipeline (Seth et al., 2024). Platforms like DeepMind’s AlphaFold have transformed structural biology, expediting the design of drugs targeting specific protein configurations (Qiu et al., 2024).

In genomics, the precision and efficiency of CRISPR-based gene editing have been significantly enhanced by AI. Tools leveraging AI predict off-target effects and guide optimal design of CRISPR guides. Moreover, machine learning models are applied to annotate genes, predict their functions, and assess the pathogenicity of genetic variants, which is pivotal for advancing gene therapy (Aljabali et al., 2024; Abbasi et al., 2025).

Synthetic biology, another frontier, benefits from AI’s role in designing synthetic gene circuits and optimising metabolic pathways. AI accelerates the engineering of cells to produce biofuels, pharmaceuticals, and other high-value products. Predictive modelling and optimisation algorithms ensure that synthetic constructs perform efficiently under diverse conditions (Amaan et al., 2024).

In microbiome research, AI has deepened our understanding of the dynamic interactions between host organisms and their microbiota. AI algorithms analyse metagenomic datasets to unravel the roles of microbial communities in health and disease. These insights drive the development of novel probiotics and therapeutic interventions, tailored to individual microbiomes (Probul et al., 2024).

Recent breakthroughs and case studies

Recent years have seen transformative breakthroughs at the interface of artificial intelligence (AI) and bioinformatics, driving substantial progress in biotechnology. One of the most notable examples is AlphaFold, developed by DeepMind, which has revolutionized protein structure prediction (Desai et al., 2024). AlphaFold's deep learning algorithms achieved remarkable accuracy in predicting protein folding, a challenge that persisted for decades in biology. This innovation has accelerated research in structural biology and has applications ranging from drug development to understanding complex biological systems (Z. Yang et al., 2023). For instance, AlphaFold is actively used in studies addressing neglected diseases such as Chagas disease and Leishmaniasis, significantly improving drug discovery pipelines by identifying viable targets (Gabaldón-Figueira et al., 2023).

Another breakthrough is the role of AI in combating the COVID-19 pandemic. AI-assisted platforms facilitated vaccine development by identifying antigen candidates and optimising vaccine design through techniques like reverse vaccinology (Olawade et al., 2024). Machine learning models have been instrumental in analysing genomic sequences of SARS-CoV-2 to predict mutations and their implications on vaccine efficacy, thereby enabling adaptive vaccine strategies (Lebatteux et al., 2024). Additionally, AI-powered molecular simulations helped prioritise drug candidates for therapeutic interventions during the pandemic (Liu et al., 2022).

Case studies highlight the success of integrating AI tools into bioinformatics pipelines for real-world biotechnology applications. In drug discovery, AI frameworks such as graph neural networks have optimized lead compound identification and protein-ligand interaction predictions, significantly shortening the time and cost associated with conventional methods (Fu, Chen, 2025). For example, AI platforms have supported the repurposing of drugs for COVID-19 treatment, streamlining the search for effective compounds (Bagabir et al., 2022).

These breakthroughs underscore the vast potential of AI in addressing challenges across diverse domains of biotechnology. The synergy between AI and bioinformatics continues to transform data-driven research, enabling precise solutions to biological problems, enhancing our ability to combat diseases, and accelerating innovation across the life sciences (Mohseni, Ghorbani, 2024).

The case studies highlighted in this section (Tab. 4) were selected based on a combination of factors, including high citation impact, scientific novelty, real-world application, and the diversity of AI methodologies employed (e.g., deep learning, graph neural networks, NLP).

Tab. 4. Summary of selected AI-driven breakthroughs in bioinformatics

Case study	Domain	AI technique	Impact /Application
AlphaFold (Desai et al., 2024)	Proteomics / Structural Biology	Deep Learning (Transformer-based)	Revolutionized protein structure prediction; used in neglected disease drug pipelines
COVID-19 Vaccine Design (Olawade et al., 2024)	Genomics / Vaccinology	Machine Learning	Enabled reverse vaccinology and antigen candidate identification

SARS-CoV-2 Mutation Prediction (Lebatteux et al., 2024)	Genomics	Machine Learning	Predicted viral mutations affecting vaccine efficacy; informed adaptive vaccine strategies
AI Drug Repurposing (Liu et al., 2022; Bagabir et al., 2022)	Drug Discovery	Molecular Simulation + ML	Prioritised compounds for COVID-19 therapy; accelerated treatment pipelines
Graph Neural Networks for Lead Optimisation (Fu, Chen, 2025)	Drug Discovery	Graph Neural Networks	Enhanced protein-ligand interaction modeling; improved hit-to-lead efficiency

When available, relevance was further supported by publication in high-impact journals or evidence of translational value such as deployment in clinical or pharmaceutical pipelines.

Challenges and limitations

AI-driven bioinformatics, while transformative, faces several significant challenges and limitations. One of the foremost technical hurdles is data quality, particularly regarding the completeness, accuracy, and representativeness of datasets used for training algorithms (Jamarani et al., 2024). High-throughput sequencing and clinical datasets often contain noise or missing values, which can bias the results and affect reproducibility (Williamson, Prybutok, 2024). Data bias is particularly problematic, with underrepresented populations in datasets leading to systemic inequalities in algorithm performance, as seen in the healthcare domain where racial, gender, and age-related biases are prevalent (Franklin et al., 2024). For example, algorithms trained on predominantly White or male datasets may fail to generalize across diverse populations, leading to disparities in AI-driven biomedical research and clinical applications (Nazer et al., 2023). Furthermore, biases in data collection, such as the overrepresentation of specific demographics or diseases, result in skewed predictions that exacerbate existing healthcare inequities (Ferrara, 2023). The lack of standardisation across bioinformatics pipelines also complicates the integration of diverse data sources, limiting interoperability and consistency in AI outcomes (Brancato et al., 2024).

Privacy and security concerns surrounding AI-driven bioinformatics represent another major ethical challenge. AI's reliance on large-scale genomic and clinical datasets raises critical issues regarding data storage, sharing, and potential misuse (Khalid et al., 2023). One of the most pressing risks is patient re-identification, where AI models can cross-reference genomic datasets with publicly available records to deduce an individual’s identity. Such risks threaten patient confidentiality and raise ethical concerns related to informed consent and data ownership (Ford et al., 2025). Additionally, AI-driven genetic discrimination has emerged as a growing concern, where individuals may face bias in employment, insurance, or healthcare decisions based on AI-predicted disease risks (Cross et al., 2024). Notably, reports have suggested that certain insurance companies have attempted to use genetic data to assess policy risks, highlighting the urgent need for stronger legal protections against discriminatory practices (Tiller et al., 2022).

Regulatory frameworks such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) aim to address genomic data privacy issues. GDPR enforces strict consent requirements, the right to data erasure, and limitations on cross-border data transfers, while HIPAA focuses on ensuring the security of protected health information (Feretzakis et al., 2024). However, these frameworks face challenges in adapting to the unique risks posed by AI-driven genetic analytics and automated decision-making. As AI governance policies evolve, ensuring transparency and fairness in genomic data applications remains an ongoing challenge for policymakers and researchers (Taddese et al., 2025). Several real-world incidents illustrate the vulnerabilities associated with genomic data misuse. In 2019, a major data breach at MyHeritage exposed the genetic information of over 92 million users, demonstrating the susceptibility of AI-driven genomic databases to cyberattacks (Arshad et al., 2021). More recently, a credential-stuffing attack on 23andMe in 2023 led to unauthorized access to genetic ancestry data, raising concerns over the security of direct-to-consumer genetic testing services (Holthouse et al., 2025). Additionally, reports indicate that AI-powered bioinformatics tools have been leveraged in state surveillance programs to analyse DNA samples for tracking ethnic populations, further complicating the ethical landscape of genomic AI applications (Khalid et al., 2023). These cases highlight the urgent need for stricter encryption protocols, enhanced data governance policies, and AI-specific regulatory frameworks to prevent data breaches and misuse.

Beyond privacy, AI ethics debates have intensified around accountability in automated decision-making. As AI-driven bioinformatics models increasingly influence biomedical research and clinical decision-making, the challenge of assigning responsibility for errors or biases has become critical (Radanliev, 2025). The lack of transparency in many AI models, often referred to as the “black-box” problem, raises concerns about interpretability, reproducibility, and trustworthiness in AI-generated insights (Pedreschi et al., 2019). These issues are particularly pressing in drug discovery, disease risk assessment, and genetic diagnostics, where incorrect predictions could lead to flawed medical decisions with significant consequences for patient health. Ensuring explainability in AI models used in bioinformatics is essential for fostering trust among researchers, clinicians, and patients (Sadeghi et al., 2024).

AI's role in biotechnology research also presents challenges in maintaining scientific integrity. The automation of hypothesis generation and data interpretation introduces risks of scientific misconduct, including AI-generated errors being misrepresented as novel findings (Z. Chen et al., 2024). In particular, reliance on AI in bioinformatics research without proper validation can lead to the propagation of fabricated or biased results, undermining the credibility of scientific discoveries (Wang et al., 2024). As AI becomes more integrated into regulatory and clinical decision-making pipelines, it is imperative to establish ethical frameworks that prioritise transparency, fairness, and accountability in AI-driven bioinformatics research (Cross et al., 2024).

Interdisciplinary barriers further hinder the seamless integration of AI into bioinformatics. Effective collaboration between biologists, clinicians, and AI experts is essential to bridge knowledge gaps and ensure that computational methods align with biological insights (Patel et al., 2024). However, the disconnect between these disciplines often results in inefficiencies, slowing down the translation of AI innovations into practical applications. Addressing these challenges requires targeted training programs that equip researchers across disciplines with the necessary computational and biological expertise (Wenger et al., 2024). Additionally, fostering interdisciplinary communication and aligning research methodologies will be critical to optimising the application of AI-driven bioinformatics in real-world contexts (Dwivedi et al., 2023).

Despite these challenges, AI continues to drive significant advancements in bioinformatics and biotechnology. Addressing data bias, enhancing genomic data security, strengthening AI accountability, and fostering interdisciplinary collaboration will be key to ensuring the ethical and effective deployment of AI-driven bioinformatics solutions. As the field progresses, proactive regulatory measures and continuous dialogue between policymakers, researchers, and industry stakeholders will be essential to balance innovation with ethical responsibility.

Future prospects

The future of AI in bioinformatics is marked by groundbreaking advances in emerging technologies and transformative approaches to data integration. Among these, quantum computing holds immense promise in revolutionising the field (Fu, Chen, 2025). Quantum computers, with their unparalleled computational power derived from phenomena such as superposition and entanglement, are set to redefine our ability to process and analyse complex biological datasets (Lu et al., 2023). Unlike classical computers, quantum systems can simultaneously evaluate multiple solutions to intricate problems, significantly accelerating tasks such as protein structure prediction and drug design (Pei, 2024). These capabilities make quantum computing particularly valuable for solving optimisation challenges in molecular docking, genomic data analysis, and systems biology modelling (Pal et al., 2023).

Despite its potential, quantum computing remains in its early stages, with several critical limitations preventing its immediate application in bioinformatics. Qubit stability is a major challenge, as qubits are highly susceptible to decoherence due to environmental noise, leading to frequent errors in computations (Memon et al., 2024). Additionally, quantum error rates remain high, requiring sophisticated error-correction techniques that significantly reduce the number of usable qubits in practical computations (Zhou et al., 2024). Hardware scalability is another barrier, as current quantum processors contain only a few hundred qubits, whereas bioinformatics applications demand stable systems with thousands – if not millions – of qubits to handle large-scale biological data processing efficiently (Gill, Buyya, 2024). While ongoing advancements in superconducting qubits, trapped ions, and topological qubits are gradually improving quantum hardware, experts predict that fully functional, fault-tolerant quantum computers capable of outperforming classical supercomputers in bioinformatics applications may not be widely available until 2035–2040 (Aasen et al., 2025). However, hybrid quantum-classical computing approaches, which leverage the strengths of both quantum and classical systems, are expected to contribute to practical applications in bioinformatics and drug discovery within the next decade (Vakili et al., 2025).

In parallel, the trend of integrating multi-omics data is expected to become a cornerstone of personalized medicine. By combining genomics, transcriptomics, proteomics, and metabolomics data, researchers can create comprehensive biological models that capture the nuances of disease mechanisms and individual variability (Tanaka, 2025). AI algorithms, particularly those leveraging deep learning, are critical for decoding these complex datasets, enabling precise biomarker identification and treatment customisation (Taherdoost, Ghofrani, 2024). Multi-omics integration also facilitates the study of epigenetic changes and their impact on health, which is essential for understanding diseases such as cancer and neurodegenerative disorders (C. Chen et al., 2023). As these technologies advance, AI-powered platforms will play a pivotal role in identifying personalized therapeutic targets and accelerating precision medicine research.

Personalised health is further supported by advances in wearable technology and remote monitoring, which provide real-time data on an individual's physiological parameters. AI-driven bioinformatics platforms can analyse this influx of data to predict health risks and recommend interventions, moving healthcare from a reactive to a proactive model (Ponnarengan et al., 2024). Moreover, the incorporation of federated learning approaches ensures data privacy by enabling decentralized analysis, which is particularly relevant for sensitive health and genomic data (Li et al., 2025). These privacy-preserving AI models allow researchers to collaborate on multi-institutional datasets without compromising patient confidentiality, facilitating secure and scalable biomedical discoveries.

Looking ahead, AI and bioinformatics are poised to drive transformative innovations across biotechnology. From revolutionising drug discovery pipelines through predictive modelling to facilitating ecological and evolutionary studies, the synergistic integration of these disciplines could redefine the boundaries of science (Niazi, 2023). As quantum computing matures over the next two decades, its integration with AI-driven bioinformatics is expected to unlock unprecedented computational capabilities, accelerating breakthroughs in molecular simulations, protein engineering, and genomic medicine (Ali, 2023). In the nearer term, continued progress in deep learning, multi-omics data integration, and secure AI models will shape the next generation of biotechnology applications (Zhang et al., 2024). Collaborative efforts between computational scientists, biologists, and material engineers will be essential in overcoming current technological barriers and realising the full potential of these advancements.

Conclusions

The integration of artificial intelligence and bioinformatics has revolutionized biotechnology by enabling the efficient processing and interpretation of biological big data. AI-driven techniques, including machine learning, deep learning, and natural language processing, have facilitated advancements in precision medicine, drug discovery, synthetic biology, and microbiome research. By automating complex analyses and uncovering hidden patterns within massive datasets, AI has accelerated biomedical discoveries and personalized healthcare innovations. However, these advancements bring significant ethical and regulatory challenges, particularly regarding genomic data privacy, algorithmic bias, and AI accountability. Frameworks such as GDPR, HIPAA, and emerging AI governance policies are essential in addressing concerns related to patient data protection and AI-driven decision-making transparency, yet ongoing regulatory evolution is necessary to keep pace with rapid technological progress.

Maximising the potential of AI in bioinformatics requires strong interdisciplinary collaboration between AI researchers, biologists, clinicians, and policymakers to ensure both scientific rigor and ethical integrity. Establishing structured AI-bioinformatics partnerships through academia-industry collaborations, open-source AI platforms, and privacy-preserving federated learning models can drive innovation while ensuring responsible AI deployment. Looking ahead, integrating AI with quantum computing and multi-omics data analysis will unlock new frontiers in genomic medicine, biomarker discovery, and evolutionary biology. By addressing current challenges and fostering cross-disciplinary cooperation, AI and bioinformatics will continue to transform biotechnology, paving the way for groundbreaking scientific advancements and real-world applications in healthcare and beyond.

Conflict of interest

The author declare no conflict of interest related to this article.

References

Aasen, D., Aghaee, M., Alam, Z., Andrzejczuk, M., Antipov, A., Astafev, M., Avilovas, L., Barzegar, A., Bauer, B., Becker, J., Bello-Rivas, J.M., Bhaskar, U., Bocharov, A., Boddapati, S., Bohn, D., Bommer, J., Bonderson, P., Borovsky, J., Bourdet, L., Boutin, S., Brown, T., Campbell, G., Casparis, L., Chakravarthi, S., Chao, R., Chapman, B.J., Chatoor, S., Wulff Christensen, A., Codd, P., Cole, W., Cooper, P., Corsetti, F., Cui, A., van Dam, W., El Dandachi, T., Daraeizadeh, S., Dumitrascu, A., Ekefjärd, A., Fallahi, S., Galletti, L., Gardner, G., Gatta, R., Gavranovic, H., Goulding, M., Govender, D., Griggio, F., Grigoryan, R., Grijalva, S., Gronin, S., Gukelberger, J., Haah, J., Hamdast, M., Hansen, E.B., Hastings, M., Heedt, S., Ho, S., Hogaboam, J., Holgaard, L., Van Hoogdalem, K., Indrapiromkul, J., Ingerslev, H., Ivancevic, L., Jablonski, S., Jensen, T., Jhoja, J., Jones, J., Kalashnikov, K., Kallaher, R., Kalra, R., Karimi, F., Karzig, T., Kimes, S., Kliuchnikov, V., Kloster, M.E., Knapp, C., Knee, D., Koski, J., Kostamo, P., Kuesel, J., Lackey, B., Laeven, T., Lai, J., de Lange, G., Larsen, T., Lee, J., Lee, K., Leum, G., Li, K., Lindemann, T., Lucas, M., Lutchyn, R., Madsen, M.H., Madulid, N., Manfra, M., Markussen, S.B., Martinez, E., Mattila, M., Mattinson, J., McNeil, R., Mei, A.R. et al. (82 additional authors not shown) (2025). *Roadmap to fault tolerant quantum computation using topological qubit arrays*. arXiv (Cornell University). https://arxiv.org/abs/2502.12252. https://doi.org/10.48550/arxiv.2502.12252

Abbasi, A.F., Asim, M.N., Dengel, A. (2025). Transitioning from wet lab to artificial intelligence: a systematic review of AI predictors in CRISPR. *Journal of Translational Medicine*, *23(1)*, 153. https://doi.org/10.1186/s12967-024-06013-w

Ali, N.H. (2023). Quantum computing and AI in healthcare: Accelerating complex biological simulations, genomic data processing, and drug discovery innovations. *World Journal of Advanced Research and Reviews*, *20(2)*, 1466–1484. https://doi.org/10.30574/wjarr.2023.20.2.2325

Aljabali, A.A., El-Tanani, M., Tambuwala, M.M. (2024). Principles of CRISPR-Cas9 technology: Advancements in genome editing and emerging trends in drug delivery. *Journal of Drug Delivery Science and Technology*, *92*, 105338. https://doi.org/10.1016/j.jddst.2024.105338

Al-Kateeb, Z.N., Abdullah, D.B. (2024). Unlocking the potential: synergizing IoT, cloud computing, and big data for a bright future. *Iraqi Journal for Computer Science and Mathematics*, *5(3)*, 25. https://doi.org/10.52866/ijcsm.2024.05.03.001

Alm, C.O. (2024). Centering Humans in Artificial Intelligence. *Proceedings of the AAAI Symposium Series*, *3(1)*, 2–3. https://doi.org/10.1609/aaaiss.v3i1.31170

Amaan, A., Prekshi, G., Prachi, S. (2024). Unlocking the transformative power of synthetic biology. *Archives of Biotechnology and Biomedicine*, *8(1)*, 009–016. https://doi.org/10.29328/journal.abb.1001039

Aradhya, S., Facio, F.M., Metz, H., Manders, T., Colavin, A., Kobayashi, Y., Nykamp, K., Johnson, B., Nussbaum, R.L. (2023). Applications of artificial intelligence in clinical laboratory genomics. *American Journal of Medical Genetics. Part C*, *Seminars in Medical Genetics*, *193(3)*, e32057. https://doi.org/10.1002/ajmg.c.32057

Arjmand, B., Hamidpour, S.K., Tayanloo-Beik, A., Goodarzi, P., Aghayan, H.R., Adibi, H., Larijani, B. (2022). Machine Learning: A New Prospect in Multi-Omics Data Analysis of Cancer. *Frontiers in Genetics*, *13*, 824451. https://doi.org/10.3389/fgene.2022.824451

Arshad, S., Arshad, J., Khan, M.M., Parkinson, S. (2021). Analysis of security and privacy challenges for DNA-genomics applications and databases. *Journal of Biomedical Informatics*, *119*, 103815. https://doi.org/10.1016/j.jbi.2021.103815

Ayyagiri, N.A., Aggarwal, N.A., Jain, N.S. (2024). Enhancing DNA Sequencing Workflow with AI-Driven Analytics. *International Journal for Research Publication and Seminars*, *15(3)*, 203–216. https://doi.org/10.36676/jrps.v15.i3.1484

Badrulhisham, F., Pogatzki-Zahn, E., Segelcke, D., Spisak, T., Vollert, J. (2023). Machine learning and artificial intelligence in neuroscience: A primer for researchers. *Brain Behavior and Immunity*, *115*, 470–479. https://doi.org/10.1016/j.bbi.2023.11.005

Bagabir, S.A., Ibrahim, N.K., Bagabir, H.A., Ateeq, R.H. (2022). Covid-19 and Artificial Intelligence: Genome sequencing, drug development and vaccine discovery. *Journal of Infection and Public Health*, *15(2)*, 289–296. https://doi.org/10.1016/j.jiph.2022.01.011

Bhatnagar, R., Sardar, S., Beheshti, M., Podichetty, J.T. (2022). How can natural language processing help model informed drug development?: a review. *JAMIA open*, *5(2)*, ooac043. https://doi.org/10.1093/jamiaopen/ooac043

Brancato, V., Esposito, G., Coppola, L., Cavaliere, C., Mirabelli, P., Scapicchio, C., Borgheresi, R., Neri, E., Salvatore, M., Aiello, M. (2024). Standardizing digital biobanks: integrating imaging, genomic, and clinical data for precision medicine. *Journal of Translational Medicine*, *22(1)*, 136. https://doi.org/10.1186/s12967-024-04891-8

Brandes, N., Goldman, G., Wang, C.H., Ye, C.J., Ntranos, V. (2023). Genome-wide prediction of disease variant effects with a deep protein language model. *Nature Genetics*, *55(9)*, 1512–1522. https://doi.org/10.1038/s41588-023-01465-0

Carini, C., Seyhan, A.A. (2024). Tribulations and future opportunities for artificial intelligence in precision medicine. *Journal of Translational Medicine*, *22(1)*, 411. https://doi.org/10.1186/s12967-024-05067-0

Chen, C., Wang, J., Pan, D., Wang, X., Xu, Y., Yan, J., Wang, L., Yang, X., Yang, M., Liu, G.P. (2023). Applications of multi-omics analysis in human diseases. *MedComm*, *4(4)*, e315. https://doi.org/10.1002/mco2.315

Chen, Q., Leaman, R., Allot, A., Luo, L., Wei, C., Yan, S., Lu, Z. (2021). Artificial Intelligence in Action: Addressing the COVID-19 Pandemic with Natural Language Processing. *Annual Review of Biomedical Data Science*, *4(1)*, 313–339. https://doi.org/10.1146/annurev-biodatasci-021821-061045

Chen, Y., Hsiao, T., Lin, C., Fann, Y.C. (2025). Unlocking precision medicine: clinical applications of integrating health records, genetics, and immunology through artificial intelligence. *Journal of Biomedical Science*, *32(1)*, 16. https://doi.org/10.1186/s12929-024-01110-w

Chen, Z., Chen, C., Yang, G., He, X., Chi, X., Zeng, Z., Chen, X. (2024). Research integrity in the era of artificial intelligence: challenges and responses. *Medicine*, *103(27)*, e38811. https://doi.org/10.1097/MD.0000000000038811

Collins, C., Dennehy, D., Conboy, K., Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Cross, J. L., Choma, M. A., Onofrey, J. A. (2024). Bias in medical AI: Implications for clinical decision-making. PLOS digital health, 3(11), e0000651. https://doi.org/10.1371/journal.pdig.0000651

Datta, S.D., Islam, M., Sobuz, M.H.R., Ahmed, S., Kar, M. (2024). Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review. *Heliyon*, *10(5)*, e26888. https://doi.org/10.1016/j.heliyon.2024.e26888

Desai, D., Kantliwala, S.V., Vybhavi, J., Ravi, R., Patel, H., Patel, J. (2024). Review of AlphaFold 3: Transformative Advances in Drug Design and Therapeutics. *Cureus*, *16(7)*, e63646. https://doi.org/10.7759/cureus.63646

Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., Baabdullah, A.M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M.A., Al-Busaidi, A.S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S.W., Davies, G.H., Davison, R.M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J.S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K.R., Latreille, P., Laumer, S., Malik, F.T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Nord, J.H., O'Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I.O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N.P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B.C., Tiwari, M.K., van der Aalst, W., Venkatesh, V., Viglia, G., Michael Wade, M., Walton, P., Wirtz, J., Ryan Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., Verykios, V.S. (2024). Privacy-Preserving techniques in generative AI and large language Models: A Narrative review. *Information*, *15(11)*, 697. https://doi.org/10.3390/info15110697

Ferrara, E. (2023). Fairness and Bias in Artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, *6(1)*, 3. https://doi.org/10.3390/sci6010003

Ford, E., Pillinger, S., Stewart, R., Jones, K., Roberts, A., Casey, A., Goddard, K., Nenadic, G. (2025). *What is the patient re-identification risk from using de-identified clinical free text data for health research?* AI and Ethics. https://doi.org/10.1007/s43681-025-00681-0

Franklin, G., Stephens, R., Piracha, M., Tiosano, S., Lehouillier, F., Koppel, R., Elkin, P.L. (2024). The Sociodemographic Biases in Machine Learning Algorithms: A Biomedical Informatics Perspective. *Life (Basel, Switzerland)*, *14(6)*, 652. https://doi.org/10.3390/life14060652

Fu, C., Chen, Q. (2025). The future of pharmaceuticals: Artificial intelligence in drug discovery and development. *Journal of Pharmaceutical Analysis*, 101248. https://doi.org/10.1016/j.jpha.2025.101248

Gabaldón-Figueira, J.C., Martinez-Peinado, N., Escabia, E., Ros-Lucas, A., Chatelain, E., Scandale, I., Gascon, J., Pinazo, M.J., Alonso-Padilla, J. (2023). State-of-the-Art in the Drug Discovery Pathway for Chagas Disease: A Framework for Drug Development and Target Validation. *Research and Reports in Tropical Medicine*, *14*, 1–19. https://doi.org/10.2147/RRTM.S415273

Gill, S. S., Buyya, R. (2024). Transforming Research with Quantum Computing. *Journal of Economy and Technology*. https://doi.org/10.1016/j.ject.2024.07.001

Holthouse, R., Owens, S., Bhunia, S. (2025). *The 23andMe Data Breach: Analyzing credential stuffing attacks, security vulnerabilities, and mitigation strategies.* arXiv (Cornell University). https://doi.org/10.48550/arxiv.2502.04303

Huo, D., Wang, X. (2024). A new era in healthcare: The integration of artificial intelligence and microbial. *Medicine in Novel Technology and Devices*, *23*, 100319. https://doi.org/10.1016/j.medntd.2024.100319

Jamarani, A., Haddadi, S., Sarvizadeh, R., Kashani, M.H., Akbari, M., Moradi, S. (2024). Big data and predictive analytics: A sytematic review of applications. *Artificial Intelligence Review*, *57(7)*, 176. https://doi.org/10.1007/s10462-024-10811-5

Jamialahmadi, H., Khalili-Tanha, G., Nazari, E., Rezaei-Tavirani, M. (2024). Artificial intelligence and bioinformatics: a journey from traditional techniques to smart approaches. *Gastroenterology and Hepatology from Bed to Bench*, *17(3)*, 241–252. https://doi.org/10.22037/ghfbb.v17i3.2977

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596(7873)*, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, 106848. https://doi.org/10.1016/j.compbiomed.2023.106848

Kumar, P., Paul, R.K., Roy, H.S., Yeasin, M., Ajit, Paul, A.K. (2024). Big Data Analysis in Computational Biology and Bioinformatics. *Methods in Molecular Biology (Clifton, N.J.)*, *2719*, 181–197. https://doi.org/10.1007/978-1-0716-3461-5_11

Larabi-Marie-Sainte, S., Aburahmah, L., Almohaini, R., Saba, T. (2019). Current Techniques for Diabetes Prediction: Review and Case Study. *Applied Sciences*, *9(21)*, 4604. https://doi.org/10.3390/app9214604

Lebatteux, D., Soudeyns, H., Boucoiran, I., Gantt, S., Diallo, A.B. (2024). Machine learning-based approach KEVOLVE efficiently identifies SARS-CoV-2 variant-specific genomic signatures. *PlosOne*, *19(1)*, e0296627. https://doi.org/10.1371/journal.pone.0296627

Li, M., Xu, P., Hu, J., Tang, Z., Yang, G. (2025). From challenges and pitfalls to recommendations and opportunities: Implementing federated learning in healthcare. *Medical Image Analysis*, *101*, 103497. https://doi.org/10.1016/j.media.2025.103497

Liao, J., Li, X., Gan, Y., Han, S., Rong, P., Wang, W., Li, W., Zhou, L. (2023). Artificial intelligence assists precision medicine in cancer treatment. *Frontiers in Oncology*, *12*, 998222. https://doi.org/10.3389/fonc.2022.998222

Liu, Z., Chen, X., Carter, W., Moruf, A., Komatsu, T. E., Pahwa, S., Chan-Tack, K., Snyder, K., Petrick, N., Cha, K., Lal-Nag, M., Hatim, Q., Thakkar, S., Lin, Y., Huang, R., Wang, D., Patterson, T. A., Tong, W. (2022). AI-powered drug repurposing for developing COVID-19 treatments. *Reference Module in Biomedical Sciences*, B978-0-12-824010-6.00005-8. https://doi.org/10.1016/B978-0-12-824010-6.00005-8

Lu, Y., Sigov, A., Ratkin, L., Ivanov, L.A., Zuo, M. (2023). Quantum computing and industrial information integration: A review. *Journal of Industrial Information Integration*, *35*, 100511. https://doi.org/10.1016/j.jii.2023.100511

Memon, Q.A., Ahmad, M.A., Pecht, M. (2024). Quantum Computing: Navigating the future of computation, challenges, and technological breakthroughs. *Quantum Reports*, *6(4)*, 627–663. https://doi.org/10.3390/quantum6040039

Mohseni, P., Ghorbani, A. (2024). Exploring the synergy of artificial intelligence in microbiology: Advancements, challenges, and future prospects. *Deleted Journal*, *1*, 100005. https://doi.org/10.1016/j.csbr.2024.100005

Munawar, H.S., Ullah, F., Qayyum, S., Shahzad, D. (2022). Big data in construction: current applications and future opportunities. *Big Data and Cognitive Computing*, *6(1)*, 18. https://doi.org/10.3390/bdcc6010018

Nazer, L.H., Zatarah, R., Waldrip, S., Ke, J.X.C., Moukheiber, M., Khanna, A.K., Hicklen, R.S., Moukheiber, L., Moukheiber, D., Ma, H., Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, *2(6)*, e0000278. https://doi.org/10.1371/journal.pdig.0000278

Ng, S., Masarone, S., Watson, D., Barnes, M.R. (2023). The benefits and pitfalls of machine learning for biomarker discovery. *Cell and Tissue Research*, *394(1)*, 17–31. https://doi.org/10.1007/s00441-023-03816-z

Niazi S.K. (2023). The Coming of Age of AI/ML in Drug Discovery, Development, Clinical Testing, and Manufacturing: The FDA Perspectives. *Drug Design, Development and Therapy*, *17*, 2691–2725. https://doi.org/10.2147/DDDT.S424991

Nouri Nojadeh, J., Bildiren Eryilmaz, N.S., Ergüder, B.I. (2023). CRISPR/Cas9 genome editing for neurodegenerative diseases. *EXCLI Journal*, *22*, 567–582. https://doi.org/10.17179/excli2023-6155

Nussinov, R., Zhang, M., Liu, Y., Jang, H. (2022). AlphaFold, Artificial Intelligence (AI), and Allostery. *Journal of Physical Chemistry. B*, *126(34)*, 6372–6383. https://doi.org/10.1021/acs.jpcb.2c04346

Olawade, D.B., Teke, J., Fapohunda, O., Weerasinghe, K., Usman, S.O., Ige, A.O., David-Olawade, A.C. (2024). Leveraging artificial intelligence in vaccine development: A narrative review. *Journal of Microbiological Methods*, *224*, 106998. https://doi.org/10.1016/j.mimet.2024.106998

Pal, S., Bhattacharya, M., Lee, S., Chakraborty, C. (2023). Quantum Computing in the Next-Generation Computational Biology Landscape: From protein folding to molecular Dynamics. *Molecular Biotechnology*, *66(2)*, 163–178. https://doi.org/10.1007/s12033-023-00765-4

Patel, A.U., Gu, Q., Esper, R., Maeser, D., Maeser, N. (2024). The crucial role of interdisciplinary conferences in advancing explainable AI in healthcare. *BioMedInformatics*, *4(2)*, 1363–1383. https://doi.org/10.3390/biomedinformatics4020075

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., Turini, F. (2019). Meaningful explanations of black box AI decision systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33(01)*, 9780–9784. https://doi.org/10.1609/aaai.v33i01.33019780

Pei, Z. (2024). Computer-aided drug discovery: From traditional simulation methods to language models and quantum computing. *Cell Reports Physical Science*, 102334. https://doi.org/10.1016/j.xcrp.2024.102334

Pereira, R., Oliveira, J., Sousa, M. (2020). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of Clinical Medicine*, *9(1)*, 132. https://doi.org/10.3390/jcm9010132

Ponnarengan, H., Rajendran, S., Khalkar, V., Devarajan, G., Kamaraj, L. (2024). Data-Driven Healthcare: The role of computational methods in medical innovation. *Computer Modeling in Engineering and Sciences*, *0(0)*, 1–10. https://doi.org/10.32604/cmes.2024.056605

Probul, N., Huang, Z., Saak, C. C., Baumbach, J., List, M. (2024). AI in microbiome-related healthcare. *Microbial Biotechnology*, *17(11)*, e70027. https://doi.org/10.1111/1751-7915.70027

Qiu, X., Li, H., Ver Steeg, G., Godzik, A. (2024). Advances in AI for Protein Structure Prediction: Implications for Cancer Drug Discovery and Development. *Biomolecules*, *14(3)*, 339. https://doi.org/10.3390/biom14030339

Quazi S. (2022). Artificial intelligence and machine learning in precision and genomic medicine. *Medical Oncology (Northwood, London, England)*, *39(8)*, 120. https://doi.org/10.1007/s12032-022-01711-1

Radanliev, P. (2025). AI Ethics: Integrating transparency, fairness, and privacy in AI development. *Applied Artificial Intelligence*, *39(1)*. https://doi.org/10.1080/08839514.2025.2463722

Rawat, R., Yadav, R. (2021). Big data: big data analysis, issues and challenges and technologies. *IOP Conference Series Materials Science and Engineering*, *1022(1)*, 012014. https://doi.org/10.1088/1757-899x/1022/1/012014

Sadeghi, Z., Alizadehsani, R., Cifci, M.A., Kausar, S., Rehman, R., Mahanta, P., Bora, P.K., Almasri, A., Alkhawaldeh, R.S., Hussain, S., Alatas, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., Pardalos, P.M. (2024). A review of Explainable Artificial Intelligence in healthcare. *Computers and Electrical Engineering*, *118*, 109370. https://doi.org/10.1016/j.compeleceng.2024.109370

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra, A.K., Das, G., Malonia, S.K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, *12(7)*, 997. https://doi.org/10.3390/biology12070997

Seth, A., Thillainadesan, G., Amemiya, Y., Nam, R. (2024). 22. Advancing personalized prostate cancer care: Utilizing miRNA profiling and machine learning for metastasis prediction. *Cancer Genetics*, *286–287*, S7–S8. https://doi.org/10.1016/j.cancergen.2024.08.024

Singh, S., Gupta, H., Sharma, P., Sahi, S. (2023). Advances in Artificial Intelligence (AI)-assisted approaches in drug screening. *Artificial Intelligence Chemistry*, *2(1)*, 100039. https://doi.org/10.1016/j.aichem.2023.100039

Spurney, R., Schwartz, M., Gobble, M., Sozzani, R., Van den Broeck, L. (2021). Spatiotemporal Gene Expression Profiling and Network Inference: A Roadmap for Analysis, Visualization, and Key Gene Identification. *Methods in Molecular Biology (Clifton, N.J.)*, *2328*, 47–65. https://doi.org/10.1007/978-1-0716-1534-8_4

Taddese, A.A., Addis, A.C., Tam, B.T. (2025). Data stewardship and curation practices in AI-based genomics and automated microscopy image analysis for high-throughput screening studies: promoting robust and ethical AI applications. *Human Genomics*, *19(1)*. https://doi.org/10.1186/s40246-025-00716-x

Taherdoost, H., Ghofrani, A. (2024). AI's role in revolutionizing personalized medicine by reshaping pharmacogenomics and drug therapy. *Intelligent Pharmacy*, *2(5)*, 643–650. https://doi.org/10.1016/j.ipha.2024.08.005

Tanaka, M. (2025). From serendipity to precision: integrating AI, Multi-Omics, and Human-Specific models for personalized neuropsychiatric care. *Biomedicines*, *13(1)*, 167. https://doi.org/10.3390/biomedicines13010167

Tang, S., Yuan, K., Chen, L. (2022). Molecular biomarkers, network biomarkers, and dynamic network biomarkers for diagnosis and prediction of rare diseases. *Fundamental Research*, *2(6)*, 894–902. https://doi.org/10.1016/j.fmre.2022.07.011

Tiller, J., Lacaze, P., Otlowski, M. (2022). The Australian moratorium on genetics and life insurance: evaluating policy compared to Parliamentary recommendations regarding genetic discrimination. *Public Health Research and Practice*, *32(4)*. https://doi.org/10.17061/phrp3242235

Umapathy, V.R., Rajinikanth B,S., Samuel Raj, R.D., Yadav, S., Munavarah, S.A., Anandapandian, P.A., Mary, A.V., Padmavathy, K.R.A. (2023). Perspective of Artificial Intelligence in Disease Diagnosis: A Review of Current and Future Endeavours in the Medical Field. *Cureus*, *15(9)*, e45684. https://doi.org/10.7759/cureus.45684

Uzoeto, H.O., Cosmas, S., Bakare, T.T., Durojaye, O.A. (2024). AlphaFold-latest: revolutionizing protein structure prediction for comprehensive biomolecular insights and therapeutic advancements. *Beni-Suef University Journal of Basic and Applied Sciences*, *13(1)*. https://doi.org/10.1186/s43088-024-00503-y

Vakili, M.G., Gorgulla, C., Snider, J., Nigam, A., Bezrukov, D., Varoli, D., Aliper, A., Polykovsky, D., Das, K.M.P., Cox, H., Lyakisheva, A., Mansob, A.H., Yao, Z., Bitar, L., Tahoulas, D., Čerina, D., Radchenko, E., Ding, X., Liu, J., Meng, F., Ren, F., Cao, Y., Stagljar, I., Aspuru-Guzik, A., Zhavoronkov, A. (2025). Quantum-computing-enhanced algorithm unveils potential KRAS inhibitors. *Nature Biotechnology*. https://doi.org/10.1038/s41587-024-02526-3

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50(D1)*, D439–D444. https://doi.org/10.1093/nar/gkab1061

Visan, A.I., Negut, I. (2024). Integrating Artificial Intelligence for Drug Discovery in the Context of Revolutionizing Drug Delivery. *Life (Basel, Switzerland)*, *14(2)*, 233. https://doi.org/10.3390/life14020233

Vitorino, R. (2024). Transforming Clinical Research: The Power of High-Throughput Omics Integration. *Proteomes*, *12(3)*, 25. https://doi.org/10.3390/proteomes12030025

Vora, L.K., Gholap, A.D., Jetha, K., Thakur, R.R.S., Solanki, H.K., Chavda, V.P. (2023). Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. *Pharmaceutics*, *15(7)*, 1916. https://doi.org/10.3390/pharmaceutics15071916

Wang, J., Cheng, Z., Yao, Q., Liu, L., Xu, D., Hu, G. (2024). Bioinformatics and biomedical informatics with ChatGPT: Year one review. *Quantitative Biology (Beijing, China)*, *12(4)*, 345–359. https://doi.org/10.1002/qub2.67

Wenger, D., Hossain, M.S., Senseman, J.R. (2024). *AI and the Impact on Journalism Education. Journalism and Mass Communication Educator*. https://doi.org/10.1177/10776958241296497

Williamson, S.M., Prybutok, V. (2024). Balancing Privacy and Progress: A Review of Privacy Challenges, Systemic Oversight, and Patient Perceptions in AI-Driven Healthcare. *Applied Sciences*, *14(2)*, 675. https://doi.org/10.3390/app14020675

Yang, C.T., Kristiani, E., Leong, Y.K., Chang, J.S. (2023). Big data and machine learning driven bioprocessing - Recent trends and critical analysis. *Bioresource Technology*, *372*, 128625. https://doi.org/10.1016/j.biortech.2023.128625

Yang, Z., Zeng, X., Zhao, Y., Chen, R. (2023). AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduction and Targeted Therapy*, *8(1)*. https://doi.org/10.1038/s41392-023-01381-z

Zhang, Y., Huang, G., Zhao, Y., Lu, X., Wang, Y., Wang, C., Guo, X., Zhao, C. (2024). *Revolutionizing crop breeding: Next-Generation Artificial intelligence and Big Data-Driven intelligent design*. Engineering. https://doi.org/10.1016/j.eng.2024.11.034

Zhao, J., Feng, Q., Wei, W.Q. (2022). Integration of Omics and Phenotypic Data for Precision Medicine. Methods in *Molecular Biology (Clifton, N.J.)*, *2486*, 19–35. https://doi.org/10.1007/978-1-0716-2265-0_2

Zhou, H., Zhao, C., Cain, M., Bluvstein, D., Duckering, C., Hu, H., Wang, S., Kubica, A., Lukin, M. D. (2024). *Algorithmic fault tolerance for fast quantum computing.* arXiv (Cornell University). https://doi.org/10.48550/arxiv.2406.17653

Zitnik, M., Li, M.M., Wells, A., Glass, K., Gysi, D.M., Krishnan, A., Murali, T.M., Radivojac, P., Roy, S., Baudot, A., Bozdag, S., Chen, D.Z., Cowen, L., Devkota, K., Gitter, A., Gosline, S.J.C., Gu, P., Guzzi, P.H., Huang, H., Jiang, M., Kesimoglu, Z.N., Koyuturk, M., Ma, J., Pico, A.R., Pržulj, N., Przytycka, T.M., Raphael, B.J., Ritz, A., Sharan, R., Shen, Y., Singh, M., Slonim, D.K., Tong, H., Yang, X.H., Yoon, B.-J., Haiyuan Yu, H.,Milenković, T. (2024). Current and future directions in network biology. *Bioinformatics Advances*, *4(1)*. https://doi.org/10.1093/bioadv/vbae099

Bioinformatyka spotyka AI: transformacja dużych zbiorów danych przełomem w biotechnologii

Streszczenie

Konwergencja bioinformatyki i sztucznej inteligencji (AI) rewolucjonizuje krajobraz biotechnologiczny, przekształcając duże zbiory danych biologicznych w praktyczne spostrzeżenia oraz przełomowe innowacje. Bioinformatyka, ze swoją zdolnością do zarządzania i analizowania ogromnych zestawów danych z genomiki, proteomiki i biologii systemów, staje przed wyzwaniami złożoności oraz wielkiej skali. Zaawansowane techniki AI – uczenie maszynowe, głębokie uczenie i przetwarzanie języka naturalnego – zapewniają bezprecedensowe narzędzia do rozszyfrowywania wzorców, tworzenia prognoz oraz napędzania automatyzacji. Ta synergia katalizuje niezwykły postęp w precyzyjnej medycynie, odkrywaniu leków, edycji genów i biologii syntetycznej, zapowiadając nową erę przełomów opartych na dużej ilości danych. W przeglądzie tym badane są podstawy, transformacyjne zastosowania i ostatnie przełomy na styku bioinformatyki i AI, a jednocześnie podejmowane są wyzwania oraz wizualizowanie przyszłości, w której interdyscyplinarna współpraca uwalnia pełny potencjał tego potężnego partnerstwa.

Słowa kluczowe: sztuczna inteligencja, duże zbiory danych, bioinformatyka, innowacje biotechnologiczne, biologia obliczeniowa

Information about the author

Kirolos Eskandar https://orcid.org/0000-0003-0085-3284

He focuses on combining medicine with cutting-edge technology to achieve tangible advances in medical innovation, in areas such as oncology, psychiatry and virology, and AI in healthcare.